# Necessary and sufficient conditions of solution uniqueness in $\ell_1$ minimization

**Hui Zhang · Wotao Yin · Lizhi Cheng**

**Abstract** This paper shows that the solutions to various convex $\ell_1$ minimization problems are *unique* if and only if a common set of conditions are satisfied. This result applies broadly to the basis pursuit model, basis pursuit denoising model, Lasso model, as well as other $\ell_1$ models that either minimize $f(Ax - b)$ or impose the constraint $f(Ax - b) \leq \sigma$, where $f$ is a strictly convex function. For these models, this paper proves that, given a solution $x^*$ and defining $I = \text{supp}(x^*)$ and $s = \text{sign}(x_I^*)$, $x^*$ is the unique solution if and only if $A_I$ has full column rank and there exists $y$ such that $A_I^T y = s$ and $|a_i^T y|_\infty < 1$ for $i \notin I$. This condition is previously known to be sufficient for the basis pursuit model to have a unique solution supported on $I$. Indeed, it is also necessary, and applies to a variety of other $\ell_1$ models. The paper also discusses ways to recognize unique solutions and verify the uniqueness conditions numerically.

**Keywords** $\ell_1$ minimization · basis pursuit · LASSO · solution uniqueness · strict complementarity

## 1 Introduction

Let $x \in \mathbb{R}^n$ be the decision variable. This paper studies the unique solutions of the $\ell_1$ minimization problems including the basis pursuit problem [8]

$$\min \|x\|_1 \quad \text{s.t. } Ax = b \tag{1a}$$

H. Zhang and L. Cheng
Department of Mathematics and Systems Science, College of Science,
National University of Defense Technology,
Changsha, Hunan, China, 410073.
E-mail: hui.zhang@rice.edu and clzcheng@nudt.edu.cn

Wotao Yin
Department of Computational and Applied Mathematics,
Rice University,
Houston, Texas, USA. 77005.
E-mail: wotao.yin@rice.edu

| 1. REPORT DATE **AUG 2012** | 2. REPORT TYPE | 3. DATES COVERED **00-00-2012 to 00-00-2012** |
|---|---|---|

| 4. TITLE AND SUBTITLE | 5a. CONTRACT NUMBER |
|---|---|
| **Necessary and sufficient conditions of solution uniqueness in l1 minimization** | 5b. GRANT NUMBER |
| | 5c. PROGRAM ELEMENT NUMBER |
| 6. AUTHOR(S) | 5d. PROJECT NUMBER |
| | 5e. TASK NUMBER |
| | 5f. WORK UNIT NUMBER |

| 7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Stanford University ,Department of Applied and Computational Mathematics,Houston,TX,77005** | 8. PERFORMING ORGANIZATION REPORT NUMBER |
|---|---|

| 9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) | 10. SPONSOR/MONITOR'S ACRONYM(S) |
|---|---|
| | 11. SPONSOR/MONITOR'S REPORT NUMBER(S) |

| 12. DISTRIBUTION/AVAILABILITY STATEMENT |
|---|
| **Approved for public release; distribution unlimited** |

| 13. SUPPLEMENTARY NOTES |
|---|

14. ABSTRACT

**This paper shows that the solutions to various convex '1 minimization problems are unique if and only if a common set of conditions are satisfied. This result applies broadly to the basis pursuit model, basis pursuit denoising model, Lasso model, as well as other '1 models that either minimize f(Ax − b) or impose the constraint f(Ax − b) ≤ σ, where f is a strictly convex function. For these models, this paper proves that, given a solution x and defining I = supp(x ) and s = sign(x I ), x is the unique solution if and only if AI has full column rank and there exists y such that $A^T_I y = s$ and |aTi y|∂ < 1 for i ∉ I. This condition is previously known to be sufficient for the basis pursuit model to have a unique solution supported on I. Indeed, it is also necessary, and applies to a variety of other '1 models. The paper also discusses ways to recognize unique solutions and verify the uniqueness conditions numerically.**

| 15. SUBJECT TERMS |
|---|

| 16. SECURITY CLASSIFICATION OF: | | | 17. LIMITATION OF ABSTRACT | 18. NUMBER OF PAGES | 19a. NAME OF RESPONSIBLE PERSON |
|---|---|---|---|---|---|
| a. REPORT **unclassified** | b. ABSTRACT **unclassified** | c. THIS PAGE **unclassified** | **Same as Report (SAR)** | **11** | |

and convex problems

$$\min f_1(Ax - b) + \lambda \|x\|_1, \tag{1b}$$

$$\min \|x\|_1, \quad \text{s.t. } f_2(Ax - b) \le \sigma, \tag{1c}$$

$$\min f_3(Ax - b), \quad \text{s.t. } \|x\|_1 \le \tau, \tag{1d}$$

where $\lambda, \sigma, \tau > 0$ are scalar parameters, $A$ is a matrix, and $f_i(x), i = 1, 2, 3$ are strictly convex functions. The Lasso problem [21] is a special case of problem (1b) or (1d) while the basis pursuit denoising problem [8] is a special case of problem (1c) all with $f_i(\cdot) = \frac{1}{2}\|\cdot\|_2^2, \ i = 1, 2, 3$.

There is a rich literature on analyzing, solving, and applying problems in forms of (1a)–(1d) in the communities of information theory, signal processing, statistics, machine learning, optimization, and so on. In many cases, problems (1a)–(1d) need to have unique solutions; when there are more than one solution, the set of solutions is a convex set including an infinite number of solutions. In compressive sensing signal recovery, having non-unique solutions means that the underlying signal can not be reliably recovered from the given data. When the basis pursuit denoising problem has a unique solution, the error between the solution and original sparse signal can be bounded by the noise size [18]. In feature selection, non-unique solutions cause ambiguity for feature identification. Even certain optimization methods and algorithms, especially those producing the solution path of (1b)–(1d) over varying parameters such as LARS [14] and parametric quadratic programming [1], require solution uniqueness; they will fail (or need special treatments) upon encountering non-unique solutions. Therefore, establishing a condition of solution uniqueness is important for both the analysis and computation of problems (1a)–(1d). Fortunately, there are various sufficient conditions guaranteeing solution uniqueness in problem (1a), such as Spark [10,2], the mutual incoherence condition [11,15], the null-space property (NSP) [12,9], the restricted isometry principle (RIP) [3], the spherical section property [24], the "RIPless" property [5], and so on. Some conditions guarantee the unique recovery of a given solution or solutions with a given set of signs; other conditions provide the guarantees for all solutions with sufficiently few nonzero entries. However, none of them is known to be both necessary and sufficient for solution uniqueness in problems (1b)–(1d). This paper shows that given a solution $x^*$ to any problem among (1a)–(1d), Condition 1 below is both necessary and sufficient for $x^*$ to be the unique solution. Hence, it is weaker than the sufficient conditions listed above.

We let $X$, $X_\lambda$, $Y_\sigma$, and $Z_\tau$ denote the sets of solutions to problems (1a)–(1d), respectively. We let $a_i$ be the $i$th column of $A$ and $x_i$ be the $i$th entry of $x$. Given an index set $I$, we frequently use $A_I$ as the submatrix of $A$ formed by its columns $a_i$, $i \in I$ and $x_I$ as the subvector of $x$ formed by entries $x_i$, $i \in I$.

Our analysis makes the following assumptions:

**Assumption 1** *Matrix $A$ has full row rank.*

**Assumption 2** *The solution sets $X$, $X_\lambda$, $Y_\sigma$, and $Z_\tau$ of problems (1a)–(1d), respectively, are nonempty.*

**Assumption 3** *In problems (1b)–(1d), functions $f_1, f_2, f_3$ are strictly convex. In addition, the constraint of problem (1d) is bounding, namely, $\tau \le \inf\{\|x\|_1 : f_3(Ax - b) = f_3^*\}$, where $f_3^* := \min_{y \in \mathbb{R}^n} f(Ay - b)$.*

Assumptions 1 and 2 are standard. If Assumption 1 does not hold and $Ax = b$ is consistent, the problems can be simplified; specifically, one can decompose $A = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix}$ and $b = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$ so that and $A_1$ has full row rank equal to rank$(A)$, and one can replace the constraints $Ax = b$ by $A_1 x = b_1$ and introduce functions $\bar{f}_i$ so that $\bar{f}_i(A_1 x - b_1) \equiv f_i(Ax - b), i = 1, 2, 3$. Assumption 2 guarantees that the solutions of problems (1a)–(1d) can be attained so the discussion of solution uniqueness makes sense. The strict convexity of $f_1, f_2, f_3$ and the restriction on $\tau$ in Assumption 3 are also fairly basic toward solution uniqueness. Strict convexity rules out

piece-wise linearity. (Note that $f_1, f_2, f_3$ can still be non-differentiable.) If the restriction on $\tau$ is removed, the solution uniqueness of problem (1d) becomes solely up to $f_3(Ax - b)$, independent of $\|x\|_1$.

For a given vector $x^*$, the following conditions on matrix $A$ is the key to solution uniqueness, and its sufficiency has been established in [16].

**Condition 1** *Under the definitions $I := \operatorname{supp}(x^*) \subseteq \{1, \ldots, n\}$ and $s := \operatorname{sign}(x_I^*)$, matrix $A \in \mathbb{R}^{m \times n}$ has the following properties:*

1. *submatrix $A_I$ has full column rank, and*
2. *there is $y \in \mathbb{R}^m$ obeying $A_I^T y = s$ and $\|A_{I^c}^T y\|_\infty < 1$.*

The main theorem of this paper asserts that Condition 1 is both necessary and sufficient to the uniqueness of solution $x^*$.

**Theorem 1 (Solution uniqueness)** *Under Assumptions 1–3, given that $x^*$ is a solution to problem* (1a), (1b), (1c), *or* (1d), *$x^*$ is the unique solution if and only if Condition 1 holds.*

In addition, combining Theorem 1 with the optimality conditions for problems (1a)–(1d), the following theorems give the necessary and sufficient conditions of *unique optimality* for those problems.

**Theorem 2 (Basis pursuit unique optimality)** *Under Assumptions 1–2, $x^* \in \mathbb{R}^n$ is the unique solution to problem* (1a) *if and only if $Ax^* = b$ and Condition 1 is satisfied.*

**Theorem 3 (Problems (1b)–(1d) unique optimality)** *Under Assumptions 1–3 and the additional assumption $f_1, f_2, f_3 \in C^1$, $x^* \in \mathbb{R}^n$ is the unique solution to problem* (1b), (1c), *or* (1d) *if and only if, respectively,*

$$\exists \, p^* \in \partial \|x^*\|_1, \ni p^* + \lambda A^T \nabla f_1(Ax^* - b) = \mathbf{0}, \tag{2a}$$

$$\begin{aligned} f(Ax^* - b) \leq \sigma \quad and \quad \exists \, p^* \in \partial \|x^*\|_1, \eta \geq 0, \\ \ni p^* + \eta A^T \nabla f_2(Ax^* - b) = \mathbf{0}, \quad or \end{aligned} \tag{2b}$$

$$\begin{aligned} \|x^*\|_1 \leq \tau \quad and \quad \exists \, p^* \in \partial \|x^*\|_1, \nu \geq 0, \\ \ni \nu p^* + A^T \nabla f_3(Ax^* - b) = \mathbf{0}, \end{aligned} \tag{2c}$$

*and in addition Condition 1 holds.*

The proofs of these theorems are given in Section 2 below.

## 1.1 Related works

Since the sufficiency is not the focus of this paper, we do not go into more details of the sufficient conditions that have been mentioned above. We just point out that several papers such as [16,6] construct the least-squares (i.e., minimal $\ell_2$-norm) solution $\bar{y}$ of $A_I^T y = s$ and establish sufficient conditions for $\|A_{I^c}^T \bar{y}\|_\infty < 1$ to hold. Next, we review the existing results on necessary conditions for the uniqueness of $\ell_1$ minimizer.

Work [4] considers problem (1a) with complex-valued quantities and $A$ equal to a down-sampled discrete Fourier operator, for which it establishes both the necessity and sufficiency of Condition 1 to the solution uniqueness of (1a). Their proof uses the Hahn-Banach separation theorem and the Parseval formula. Work [23] lets the entries of matrix $A$ and vector $x$ in problem (1a) have complex values and gives a sufficient

condition for its solution uniqueness. In regularization theory, Condition 1 is used to derive linear error bounds under the name of range or source conditions in [18], which shows the necessity and sufficiency of Condition 1 for solution uniqueness of (1a) in a Hilbert-space setting. More recently, [13] constructs the set

$$\mathbb{F} = \{x : \|A_{J^c}^T A_J (A_J^T A_J)^{-1}\text{sign}(x_J)\|_\infty < 1 \text{ and } \text{rank}(A_J) = |J|\},$$

where $J = \text{supp}(x)$, and then states that the set of vectors that can be recovered by problem (1a) is exactly characterized by the closure of $\mathbb{F}$ if the measurement matrix $A$ satisfies the so-called *general position* (GP) condition, namely, for any sign vector $s \in \{-1, 1\}^n$, the set of columns $\{A^i\}$ of $A \in \mathbb{R}^{m \times n}$ satisfying that any $k$-dimensional affine subspace of $\mathbb{R}^m$, $k < m$, contains at most $k + 1$ points from the set $\{s_i A^i\}$. This paper claims that the result holds without the GP condition.

To our knowledge, there are few conditions addressing the solution uniqueness of problems (1b)–(1d). The following conditions in [16], [17] are sufficient for $x^*$ to the unique minimizer of (1b) for $f_1(\cdot) = \frac{1}{2}\|\cdot\|_2^2$:

$$A_I^T(b - A_I x_I^*) = \lambda \cdot \text{sign}(x_I^*), \tag{3a}$$

$$\|A_{I^c}^T(b - A_I x_I^*)\|_\infty < \lambda, \tag{3b}$$

$$A_I \text{ has full column rank.} \tag{3c}$$

However, they are not necessary as demonstrated by the following example. Let

$$A = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 2 & -2 \end{bmatrix}, \ b = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \ \lambda = 1 \tag{4}$$

and consider solving the Lasso problem, which is a special case of problem (1b):

$$\min \frac{1}{2}\|Ax - b\|_2^2 + \lambda\|x\|_1. \tag{5}$$

One gets the *unique* solution $x^* = [0 \ 1/4 \ 0]^T$ and $I = \text{supp}(x^*) = \{2\}$. However, the inequality in condition (3b) holds with equality. In general, conditions (3) becomes necessary in case $A_I$ happens to be a full rank square matrix. This assumption, however, does not apply to a sparse solution $x^*$. Nevertheless, we summarize the result in the following corollary, whose proof is given at the end of Section 2.

**Corollary 1** *If $x^*$ is the unique minimizer of problem* (1b) *with $f_1(\cdot) = \frac{1}{2}\|\cdot\|_2^2$ and if $A_I$, where $I = \text{supp}(x^*)$, is a square matrix with full rank, then condition* (3) *holds.*

Very recently, work [22] investigates the solution uniqueness of (5) and presents the following result.

**Theorem 4 ([22])** *Let $x^*$ be a solution of* (5) *and $J := \{i : |a_i^T(b - Ax^*)| = \lambda\}$. If submatrix $A_J$ is full column rank, then $x^*$ is unique. Conversely, for* almost *every $b \in \mathbb{R}^m$, if $x^*$ is unique, then $A_J$ is full column rank.*

In Theorem 4, the necessity part "for almost every $b$" is new. Indeed, it is not for every $b$. An example is given in (4) with a unique solution $x^*$ and $J = \{1, 2, 3\}$, but $A_J$ does not full column rank. On the other hand, we can figure out a special case in which the full column rankness of $A_J$ becomes necessary for all $b$ in the following corollary, whose proof is given at the end of Section 2.

**Corollary 2** *Let $x^*$ be a solution of problem* (5) *and define $I := \text{supp}(x^*)$ and $J := \{i : |a_i^T(b - Ax^*)| = \lambda\}$. If $|J| = |I| + 1$, then $x^*$ is the unique solution if and only if $A_J$ has full column rank.*

## 2 Proofs of Theorems 1–3

We establish Theorem 1 in three steps. Our first step proves the theorem for problem (1a) only. Since the only difference between this part and Theorem 2 is the conditions $Ax^* = b$, we prove Theorem 2 first. In the second step, for problems (1b)–(1d), we show that both $\|x\|_1$ and $Ax - b$ are constant for $x$ over the solution sets $X_\lambda, Y_\sigma, Z_\tau$, respectively. Finally, we prove Theorem 1 for problems (1b)–(1d).

*Proof (Proof of Theorem 2)* We frequently use the notions $I = \text{supp}(x^*)$ and $s = \text{sign}(x_I^*)$ below.

"$\Longleftarrow$". This part has been shown in [16, 23]. For completeness, we give a proof. Let $y$ satisfy Condition 1, part 2, and let $x \in \mathbb{R}^n$ be an arbitrary vector satisfying $Ax = b$ and $x \neq x^*$. We shall show $\|x^*\|_1 < \|x\|_1$.

Since $A_I$ has full column rank and $x \neq x^*$, we have $\text{supp}(x) \neq I$; otherwise from $A_I x_I^* = b = A_I x_I$, we would get $x_I^* = x_I$ and thus the contradiction $x^* = x$.

From $\text{supp}(x) \neq I$, we get $b^T y < \|x\|_1$. To see this, let $J := \text{supp}(x) \setminus I$, which is a non-empty subset of $I^c$. From Condition 1, we have $\|A_I^T y\|_\infty = 1$ and $\|A_J^T y\|_\infty < 1$, and thus

$$\langle x_I, A_I^T y \rangle \leq \|x_I\|_1 \cdot \|A_I^T y\|_\infty \leq \|x_I\|_1,$$
$$\langle x_J, A_J^T y \rangle \leq \|x_J\|_1 \cdot \|A_J^T y\|_\infty < \|x_J\|_1,$$

(the last inequality is "$<$" not "$\leq$") which lead to

$$b^T y = \langle x, A^T y \rangle = \langle x_I, A_I^T y \rangle + \langle x_J, A_J^T y \rangle$$
$$< \|x_I\|_1 + \|x_J\|_1 = \|x\|_1.$$

On the other hand, we have

$$\|x^*\|_1 = \langle x_I^*, \text{sign}(x_I^*) \rangle = \langle x_I^*, A_I^T y \rangle = \langle A_I x_I^*, y \rangle = b^T y$$

and thus $\|x^*\|_1 = b^T y < \|x\|_1$.

"$\Longrightarrow$". Assume that $x^*$ is the unique solution to (1a). Obviously, $Ax^* = b$.

It is easy to obtain Condition 1, part 1. Suppose it does not hold. Then, $A_I$ has a nontrivial null space, and perturbing $x_I^*$ along the null space will change the objective $\|x_I^*\|_1 = s^T x_I^*$ while maintaining $A_I x_I^* = b$; hence, this perturbing breaks the unique optimality of $x^*$. (In more details, there exists a *nonzero* vector $d \in \mathbb{R}^n$ such that $A_I d_I = \mathbf{0}$ and $d_{I^c} = \mathbf{0}$. For any scalar $\alpha$ near zero, we have $\text{sign}(x_I^* + \alpha d_I) = \text{sign}(x_I^*) = s$ and thus $\|x^* + \alpha d\|_1 = s^T(x_I^* + \alpha d_I) = s^T x_I^* + \alpha(s^T d_I) = \|x^*\|_1 + \alpha(s^T d_I)$. Since $x^*$ is the unique solution, we must have $\|x_I^* + \alpha d\|_1 > \|x^*\|_1$ or, equivalently, $\alpha(s^T d_I) > 0$ whenever $\alpha \neq 0$. This is impossible as we can perturb $\alpha$ around 0 both ways.)

It remains to construct a vector $y$ for Condition 1, part 2. Our construction is based on the strong convexity relation between a linear program (called the primal problem) and its dual problem, namely, if one problem has a solution, so does the other, and the two solutions must give the same objective value. (For the interested reader, this result follows from the Hahn-Banach separation theorem, also from the theorem of alternatives [7]. Alternatively, it can be obtain constructively via the Simplex method; specifically, whenever a primal solution exists, the Simplex method terminates in a finite number of steps with a primal-dual solution pair.)

The strong duality relation holds between (1a) and its dual problem

$$\max_{p \in \mathbb{R}^m} b^T p \quad \text{s.t. } \|A^T p\|_\infty \leq 1 \tag{6}$$

because (1a) and (6), as a primal-dual pair, are equivalent to the primal-dual linear programs

$$\min_{u,v \in \mathbb{R}^n} \mathbf{1}^T u + \mathbf{1}^T v \quad \text{s.t. } Au - Av = b, \ u \geq 0, \ v \geq 0, \tag{7a}$$

$$\max_{q \in \mathbb{R}^m} b^T q \quad \text{s.t. } -\mathbf{1} \leq A^T q \leq \mathbf{1}, \tag{7b}$$

respectively, where the strong duality relation holds between (7a) and (7b). By "equivalent", we mean that one can obtain solutions from each other by the rules:

$$\text{given } u^*, v^*, \text{ obtain } x^* = u^* - v^*$$
$$\text{given } x^*, \quad \text{ obtain } u^* = \max(x^*, \mathbf{0}), \ v^* = \max(-x^*, \mathbf{0}),$$
$$\text{given } q^*, \quad \text{ obtain } p^* = q^*,$$
$$\text{given } p^*, \quad \text{ obtain } q^* = p^*.$$

Therefore, since (1a) has solution $x^*$, there exists a solution $y^*$ to (6), which satisfies $\|x^*\|_1 = b^T y^*$ and $\|A^T y^*\|_\infty \leq 1$. (One can obtain such $y^*$ from the Hahn-Banach separation theorem or the theorem of alternatives rather directly.) However, $y^*$ may *not* obey $\|A_{I^c}^T y^*\|_\infty < 1$. We shall perturb $y^*$ so that $\|A_{I^c}^T y^*\|_\infty < 1$.

To prepare for the perturbation, we let $L := \{i \in I^c : a_i^T y^* = -1\}$ and $U := \{i \in I^c : a_i^T y^* = 1\}$. Our goal is to perturb $y^*$ so that $-1 < a_i^T y^* < 1$ for $i \in L \cup U$ and $y^*$ remains optimal to (6). To this end, consider for a fixed $\alpha > 0$ and $t := \|x^*\|_1$, the linear program

$$\min_{x \in \mathbb{R}^n} \sum_{i \in L} \alpha x_i - \sum_{i \in U} \alpha x_i, \quad \text{s.t. } Ax = b, \ \|x\|_1 \leq t. \tag{8}$$

Since $x^*$ is the unique solution to (1a), it is the unique feasible solution to problem (8), so problem (8) has the optimal objective value $\sum_{i \in U} \alpha x_i^* - \sum_{i \in L} \alpha x_i^* = 0$. By setting up equivalent linear programs like what has been done for (1a) and (6) above, the strong duality relation holds between (8) and its dual problem

$$\max_{p \in \mathbb{R}^m, q \in \mathbb{R}} b^T p - tq, \quad \text{s.t. } \|A^T p - \alpha r\|_\infty \leq q, \ q \geq 0, \tag{9}$$

where $r \in \mathbb{R}^n$ is given by

$$r_i = \begin{cases} 1, & i \in L, \\ -1, & i \in U, \\ 0, & \text{otherwise.} \end{cases}$$

Therefore, (9) has a solution $(p^*, q^*)$ satisfying $b^T p^* - tq^* = 0$.

According to the last constraint of (9), we have $q^* \geq 0$, which we split into two cases: $q^* = 0$ and $q^* > 0$.

i) If $q^* = 0$, we have $A^T p^* = \alpha r$ and $b^T p^* = 0$.
ii) If $q^* > 0$, we let $r^* := p^*/q^*$, which satisfies $b^T r^* = t = \|x^*\|_1$ and $\|A^T r^* - \frac{\alpha}{q^*} r\|_\infty \leq 1$, or equivalently, $-\mathbf{1} + \frac{\alpha}{q^*} r \leq A^T r^* \leq \mathbf{1} + \frac{\alpha}{q^*} r$.

Now we perturb $y^*$. Solve (9) with a sufficiently small $\alpha > 0$ and obtain a solution $(p^*, q^*)$. If case i) occurs, we let $y^* \leftarrow y^* + p^*$; otherwise, we let $y^* \leftarrow \frac{1}{2}(y^* + r^*)$. In both cases,

- $b^T y^*$ is unchanged, still equal to $\|x^*\|_1$;
- $-1 < a_i^T y^* < 1$ holds for $i \in L \cup U$ after the perturbation;
- for each $i \notin L \cup U$, if $a_j^T y^* \in [-1, 1]$ or $a_j^T y^* \in (-1, 1)$ holds before the perturbation, the same holds after the perturbation;

Therefore, after the perturbation, $y^*$ satisfies:

1) $b^T y^* = \|x^*\|_1$,

2) $\|A_I^T y^*\|_\infty \le 1$, and

3) $\|A_{I^c}^T y^*\|_\infty < 1$.

From 1) and 2) it follows

4) $A_I^T y = \text{sign}(x_I^*)$

since $\|x_I^*\|_1 = \|x^*\|_1 = b^T y^* = \langle A_I x_I^*, y^* \rangle = \langle x_I^*, A_I^T y^* \rangle \le \|x_I^*\|_1 \|A^T y^*\|_\infty \le \|x_I^*\|_1$ and thus $\langle x_I^*, A_I^T y^* \rangle = \|x_I^*\|_1$, which dictates 4). From 3) and 4), Condition 1, part 2, holds with $y = y^*$.

*Proof (Proof of Theorem 1 for problem* (1a)*)* The above proof for Theorem 2 also serves the proof of Theorem 1 for problem (1a) since $Ax^* = b$ is involved only in the optimality part, not the uniqueness part.

Next, we show that $Ax - b$ is constant for $x$ over $X_\lambda, Y_\sigma, Z_\tau$, and we first prove a simple lemma.

**Lemma 1** *Let $f$ be a strictly convex function. If $f(Ax - b) + \|x\|_1$ is constant on a convex set $S$, then both $Ax - b$ and $\|x\|_1$ are constant on $S$.*

*Proof* It suffices to prove the case where $S$ has more one point. Let $x_1$ and $x_2$ be any two different points in $S$. Consider the line segment $L$ connecting $x_1$ and $x_2$. Since $X$ is convex, we have $L \subset X$ and that $f(Ax - b) + \|x\|_1$ is constant on $L$. On one hand, $\|x\|_1$ is piece-wise linear over $L$; on the other hand, the strict convexity of $f$ makes it impossible for $f(Ax-b)$ to be piece-wise linear over $L$ unless $Ax_1 - b = Ax_2 - b$. Hence, we have $Ax_1 - b = Ax_2 - b$ and thus $f(Ax_1 - b) = f(Ax_2 - b)$, from which it follows $\|x_1\|_1 = \|x_2\|_1$. Since $x_1$ and $x_2$ are arbitrary two points in $S$, the lemma is proved.

With Lemma 1 we can show

**Lemma 2** *Under Assumptions 2 and 3, the following statements for problems* (1b)*–*(1d) *hold*

*1) $X_\lambda, Y_\sigma$ and $Z_\tau$ are convex;*

*2) In problem* (1b)*, $Ax - b$ and $\|x\|_1$ are constant for all $x \in X_\lambda$;*

*3) Part 2) holds for problem* (1c) *and $Y_\sigma$;*

*4) Part 2) holds for problem* (1d) *and $Z_\tau$.*

*Proof* Assumption 2 makes sure that $X_\lambda, Y_\sigma, Z_\tau$ are all non-empty.

1) As a well-known result, the set of solutions of a convex program is convex.

2) Since $f_1(Ax - b) + \lambda \|x\|_1$ is constant over $x \in X_\lambda$ and $f_1$ is strictly convex by Assumption 3, the result follows directly from Lemma 1.

3) If $\mathbf{0} \in Y_\sigma$, then the optimal objective is $\|\mathbf{0}\|_1 = 0$; hence, $Y_\lambda = \{\mathbf{0}\}$ and the results hold trivially. Suppose $\mathbf{0} \notin Y_\sigma$. Since the optimal objective $\|x\|_1$ is constant for all $x \in Y_\sigma$ and $f_2$ is strictly convex by Assumption 3, to prove this part in light of Lemma 1, we shall show $f_2(Ax - b) = \sigma$ for all $x \in Y_\sigma$.

Assume that there is $\hat{x} \in Y_\sigma$ such that $f_2(A\hat{x} - b) < \sigma$. Since $f_2(Ax - b)$ is convex and thus continuous in $x$, there exists a non-empty ball $\mathcal{B}$ centered at $\hat{x}$ with a sufficiently small radius $\rho > 0$ so that $f_2(A\bar{x} - b) < \sigma$ for all $\bar{x} \in \mathcal{B}$. Let $\alpha = \min\{\frac{\rho}{2 \cdot \|\hat{x}\|_2}, \frac{1}{2}\} \in (0, 1)$. We have $(1 - \alpha)\hat{x} \in \mathcal{B}$ and $\|(1 - \alpha)\hat{x}\|_1 = (1 - \alpha)\|\hat{x}\|_1 < \|\hat{x}\|_1$, so $(1 - \alpha)\hat{x}$ is both feasible and achieving an objective value lower than the optimal one. Contradiction.

4) By Assumption 3, we have $\|x\|_1 = \tau$ for all $x \in Z_\tau$; otherwise, there exists $\bar{x} \in Z_\tau$ such that $\tau > f_3(A\bar{x} - b) \ge \inf\{\|x\|_1 : f_3(Ax - b) = f_3^*\}$, contradicting Assumption 3. Since the optimal objective $f_3(Ax - b)$ is constant for all $x \in Z_\tau$ and $f_3$ is strictly convex by Assumption 3, the result follows from Lemma 1.

*Proof (Proof of Theorem 1 for problems* (1b)–(1d)*)* This proof exploits Lemma 2. Since the results of Lemma 2 are are identical for problems (1b)–(1d), we present the proof for problem (1b). The proofs for the other two problems are similar.

From Assumption 3, $X_\lambda$ is nonempty so we pick $x^* \in X_\lambda$. Let $b^* = Ax^*$, which is independent of the choice of $x^*$ according to Lemma 2. We introduce the linear program

$$\min \|x\|_1, \quad \text{s.t. } Ax = b^*, \tag{10}$$

and let $X^*$ denote its solution set.

Now, we show that $X_\lambda = X^*$. Since $Ax = Ax^*$ and $\|x\|_1 = \|x^*\|_1$ for all $x \in X_\lambda$ and conversely any $x$ obeying $Ax = Ax^*$ and $\|x\|_1 = \|x^*\|_1$ belongs to $X_\lambda$, it is suffices to show that $\|x\|_1 = \|x^*\|_1$ for any $x \in X^*$. Assuming this does *not* hold, then since problem (10) has $x^*$ as a feasible solution and has a finite objective, we have a nonempty $X^*$ and there exists $\bar{x} \in X^*$ satisfying $\|\bar{x}\|_1 < \|x^*\|_1$. But, $f(A\bar{x} - b) = f(b^* - b) = f(Ax^* - b)$ and $\|\bar{x}\|_1 < \|x^*\|_1$ mean that $\bar{x}$ is a strictly better solution to problem (1b) than $x^*$, contradicting the assumption $x^* \in X_\lambda$.

Since $X_\lambda = X^*$, $x^*$ is the unique solution to problem (1b) if and only if it is the same to problem (10). Since problem (10) is in the same form of problem (1a), applying the part of Theorem 1 for problem (1a), which is already proved, we conclude that $x^*$ is the unique solution to problem (1b) if and only if Condition 1 holds.

*Proof (Proof of Theorem 3)* The proof above also serves the proof for Theorem 3 since (2a)–(2c) are the optimality conditions of $x^*$ to problems (1b)–(1d), respectively, and furthermore, given the optimality of $x^*$, Condition 1 is the necessary and sufficient condition for the uniqueness of $x^*$.

*Remark 1* For problems (1b)–(1d), the uniqueness of a given solution $x^* \neq \mathbf{0}$ is also equivalent to a condition that is slightly simpler than Condition 1. To present the condition, consider the first-order optimality conditions (the KKT conditions) (2a)–(2c) of $x^*$ to problems (1b)–(1d), respectively, Given $x^* \neq \mathbf{0}$, $\eta$ and $\nu$ can be computed. From $p^* \neq \mathbf{0}$ it follows that $\eta > 0$. Moreover, $\nu = 0$ if and only if $A^T \nabla f_3(Ax^* - b) = \mathbf{0}$. The condition below for the case $\nu = 0$ in problem (1d) reduces to Condition 1. Define

$$\begin{aligned} P_1 &= \{i : |\lambda a_i^T \nabla f_1(Ax^* - b)| = 1\}, \\ P_2 &= \{i : |\eta a_i^T \nabla f_2(Ax^* - b)| = 1\}, \\ P_3 &= \{i : |a_i^T \nabla f_2(Ax^* - b)| = \nu\}. \end{aligned}$$

By the definitions of $\partial \|x^*\|_1$ and $P_i$, we have $\text{supp}(x^*) \subseteq P_i$, $i = 1, 2, 3$.

**Condition 2** *Under the definitions $I := \text{supp}(x^*) \subseteq P_i$ and $s := \text{sign}(x_I^*)$, matrix $A_{P_i} \in \mathbb{R}^{m \times |P_i|}$ obeys*

*1. submatrix $A_I$ has full column rank, and*
*2. there exists $y \in \mathbb{R}^m$ such that $A_I^T y = s$ and $\|A_{P_i \setminus I}^T y\|_\infty < 1$.*

*Compared to Condition 1, Condition 2 only checks the submatrix $A_{P_i}$ but not the full matrix $A$.*

It is not difficult to show that the linear programs

$$\min \|x\|_1, \quad \text{s.t. } (A_{P_i})x = b^*,$$

for $i = 1, 2, 3$, have the solution sets that are equal to $X_\lambda, Y_\sigma, Z_\tau$, respectively. From this argument, we have

**Theorem 5** *Under Assumptions 1–3 and the additional condition that $f_1, f_2, f_3 \in C^1$, given that $x^* \neq \mathbf{0}$ is a solution to problem* (1b)*,* (1c)*, or* (1d)*, $x^*$ is the unique solution if and only if Condition 2 holds for $i = 1$, 2, or 3, respectively.*

*Proof (Proof of Corollary 1)* From Theorem 3, if $x^*$ is the unique minimizer of problem (1b) with $f_1(\cdot) = \frac{1}{2}\|\cdot\|_2^2$, then Condition 1 holds, so there must exist a vector $y$ such that $A_I^T y = s$ and $\|A_{I^c}^T y\|_\infty < 1$. Combining with (3a), we have

$$\lambda A_I^T y = \lambda s = A_I^T(b - A_I x^*).$$

Since $A_I$ is a full rank square matrix, we get $y = \frac{1}{\lambda}(b - A_I x^*)$. Substituting this formula to $\|A_{I^c}^T y\|_\infty < 1$, we obtain condition (3).

*Proof (Proof of Corollary 2)* The sufficiency part follows from Theorem 4. We shall show the necessity part, namely, if $x^*$ is the unique solution, then $A_J$ has full column rank. Following the assumption $|J| = |I| + 1$, we let $\{i_0\} = J \backslash I$. Since $x^*$ is the unique solution, from Theorem 3, we know that $A_I$ has full column rank. Hence, if $A_J$ does *not* have full column rank, then we can have $a_{i_0} = A_I \beta$ for some $\beta \in \mathbb{R}^{|I|}$. From Theorem 3, if $x^*$ is the unique minimizer, then Condition 1 holds, and in particular, there must exist a vector $y$ such that $A_I^T y = s$ and $\|A_{I^c}^T y\|_\infty < 1$. Now, on one hand, as $i_0 \in I^c$, we get $1 > |a_{i_0}^T y| = |\beta^T A_I^T y| = |\beta^T s|$; on the other hand, as $i_0 \in J$, we also have $|a_{i_0}^T(b - Ax^*)| = \lambda$, which implies $1 = \frac{1}{\lambda}|a_{i_0}^T(b - Ax^*)| = \frac{1}{\lambda}|\beta^T A_I^T(b - Ax^*)| = |\beta^T s|$, where the last equality follows from (2a) (which includes (3a)). Contradiction.

## 3 Recognizing and verifying unique solutions

Applying Theorem 1, we can recognize the uniqueness of a given solution $x^*$ to problem (1a) given a dual solution $y^*$ (a solution to problem (6)). Specifically, let $J := \{i : |a_i^T y^*| = 1\}$, and if $A_J$ has full column rank and $\text{supp}(x^*) = J$, then according to Theorem 1, $x^*$ is the unique solution to (1a). The converse is not true since there generally exist multiple dual solutions with different $J$. The key is to find the one with the smallest $J$. Several linear programming interior point methods (see [19] for example) return the dual solution $y^*$ with the smallest $J$, so if either $A_J$ is column-rank deficient or $\text{supp}(x^*) \neq J$, then $x^*$ is surely non-unique.

**Corollary 3** *Under Assumption 1, given a pair of primal-dual solutions $(x^*, y^*)$ to problem (1a), let $J := \{i : |a_i^T y^*| = 1\}$. Then, $x^*$ is the unique solution to (1a) if $A_J$ has full column rank and $\text{supp}(x^*) = J$. In addition, if $y^*$ is obtained by a linear programming interior-point algorithm, the converse also holds.*

Similar results will also hold for problems (1b)–(1d) if a dual solution $y^*$ to (10) is available.

One can also directly verify Condition 1. Given a matrix $A \in \mathbb{R}^{m \times n}$, a set of its columns indexed by $I$, and a sign pattern $s = \{-1, 1\}^{|I|}$, we mention two approaches to verify Condition 1. Checking whether $A_I$ has full column rank is straightforward.

To check part 2 of Condition 1, the first approach is to follow the proof of Theorem 1. Note that Condition 1 depends only on $A$, $I$, and $s$, independent of $x^*$. Therefore, construct an *arbitrary* $x^*$ such that $\text{supp}(x^*) = I$ and $\text{sign}(x_I^*) = s$ and let $b = Ax^*$. Solve problem (6) and let $y^*$ be its solution. If $y^*$ satisfies part 2 of Condition 1, we are done; otherwise, define $L$, $U$, and $t$ by $x^*$ as in the proof, pick a small $\bar{\alpha} > 0$, and solve program (9) parametrically in $\alpha \in [0, \bar{\alpha}]$. The solution is piece-wise linear in $\alpha$ (it is possible that the solution does not exist over certain intervals of $\alpha$). Then check if there is a perturbation to $y^*$ so that $y^*$ satisfies part 2 of Condition 1.

In the second approach to check part 2 of Condition 1, one can solve the convex program

$$\min_{y \in \mathbb{R}^m} -\sum_{i \in I^c} \log(1 - a_i^T y) + \log(1 + a_i^T y),$$
$$\text{s.t. } A_I^T y = s. \tag{11}$$

Since $a_i^T y \to 1$ or $a_i^T y \to -1$ will infinitely increase the objective, (11) will return a solution satisfying Condition 1, part 2, as long as a solution exists. In fact, any feasible solution to (11) with a finite objective satisfies Condition 1, part 2. To find a feasible solution, one can apply the augmented Lagrangian method, which does not require $A_I^T y = s$ to hold at the initial point (which must still satisfy $|a_i^T y| < 1$ for all $i \in I^c$), or one can consider applying the alternating direction method of multipliers (ADMM) to the equivalent problem

$$\min_{y,z} - \sum_{i \in I^c} \log(1 - z_i) + \log(1 + z_i),$$
$$\text{s.t. } A_I^T y = s, \ z - A_{I^c}^T y = \mathbf{0}. \tag{12}$$

One can start ADMM from the origin, and the two subproblems of ADMM have closed-form solutions; in particular, the $z$-subproblem is separable in $z_i$'s and reduces to finding the zeros of 3-order polynomials in $z_i$, $i \in I^c$. If (12) has a solution, ADMM will find one; otherwise, it will diverge.

It is worth mentioning that one can use alternating projection in [20] to generate test instances that fulfill Condition 1.

## 4 Conclusions

This paper shows that Condition 1, which is previously known to be sufficient for the solution uniqueness of the basis pursuit model, is also necessary. Moreover, the condition applies to various $\ell_1$ minimization models. The result essentially follows from the fact that a pair of feasible primal-dual programs have strict complementary solutions. The result also sheds light on numerically recognizing unique solutions and verifying solution uniqueness.

## References

1. M. J. Best, and R. R. Grauer, "Sensitivity analysis for mean-variance portfolio problems," Management Science, vol. 37, no. 8, pp. 980-990, 1991.
2. A. M. Bruckstein, D. L. Donoho, and M. Elad, "From Sparse Solutions of Systems of Equations to Sparse Modeling of Signals and Images," SIAM Review, vol. 51, no. 1, pp. 34-81, 2009.
3. E. J. Candès and T. Tao, "Decoding by linear programming," IEEE Trans. Inform. Theory, vol. 51, pp. 4203-4215, 2005.
4. E. J. Candès, J. Romberg and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," IEEE Trans. Inform. Theory, vol. 52, no. 2, pp. 489-509, 2006.
5. E. J. Candès and Y. Plan, "A probabilistic and RIPless theory of compressed sensing," IEEE Trans. Inform. Theory, vol. 57, no. 11, pp. 7235-7254, 2011.
6. E. J. Candès and B. Recht, "Simple bounds for low-complexity model reconstruction," To appear in Math. Program., 2012.
7. G. B. Dantzig, Linear Programming and Extensions, Princeton, 1963.
8. S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," SIAM J. Sci. Comput., vol. 20, pp. 33-61, 1999.
9. A. Cohen, W. Dahmen, and R. DeVore, "Compressed sensing and best k-term approximation," J. Amer. Math. Soc., vol. 20, pp. 211-231, 2009.
10. D. L. Donoho and M. Elad, "Optimally sparse representation in general(non-orthogonal) dictionaries via $\ell_1$ minimization," Proc. Natl. Acad. Sci., vol. 100, no. 5, pp. 2197-2202, 2003.
11. D. L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," IEEE Trans. Inform. Theory, vol. 47, no. 7, pp. 2845-2862, 2001.
12. D. L. Donoho, "Compressed sensing," IEEE Trans. Inform. Theory, vol. 52, no. 4, pp.1289-1306, 2006.

13. C. Dossal, "A necessary and sufficient condition for exact sparse recovery by $\ell_1$ minimization," C. R. Acad. Sci. Paris, Ser. I, vol. 350, pp. 117-120, 2012.

14. B. Efron, I. Johnstone, T. Hastie, and R. Tibshirani, "Least angle regression," Ann. of Statist., vol. 32, pp. 407-499, 2004.

15. M. Elad and A. M. Bruckstein, "A generalized uncertainty principle and sparse representation in pairs of bases," IEEE Trans. Inform. Theory, vol. 48, no. 9, pp. 2558-2567, 2002.

16. J. J. Fuchs, "On sparse representations in arbitrary redundant bases," IEEE Trans. Inform. Theory, vol. 50, no. 6, pp.1341-1344, 2004.

17. J. J. Fuchs, "Recovery of Exact Sparse Representations in the Presence of Bounded Noise," IEEE Trans. Inform. Theory, vol. 51, no. 10, pp. 3601-3608, 2005.

18. M. Grasmair, M. Haltmeier, and O. Scherzer, "Necessary and Sufficient Conditions for Linear Convergence of $\ell_1$-Regularization," Comm. Pure Appl. Math., vol. 64, no. 2, pp.161-182, 2011.

19. O. Güler and Y. Ye, "Convergence behavior of interior-point algorithms," Math. Program., vol. 60, pp. 215-228, 1993.

20. D. Lorenz, "Constructing test instances for Basis Pursuit Denoising," arXiv.org/abs/1103.2897, 2011.

21. R. Tibshirani, "Regression shrinkage and selection via the lasso," J. R. Statist. Soc. B, vol. 58, no. 1, pp. 267-288, 1996.

22. R. J. Tibshirani, "The Lasso problem and uniqueness," arXiv:1206.0313v1, 2012.

23. J. A. Tropp, "Recovery of short, complex linear combinations via $\ell_1$ minimization," IEEE Trans. Inform. Theory, vol. 51, no. 4, pp. 1568-1570, 2005.

24. Y. Zhang, Theory of compressive sensing via $\ell_1$-minimization: a non-RIP analysis and extensions, Rice University, Houston, TX, Tech. Rep., 2008.